

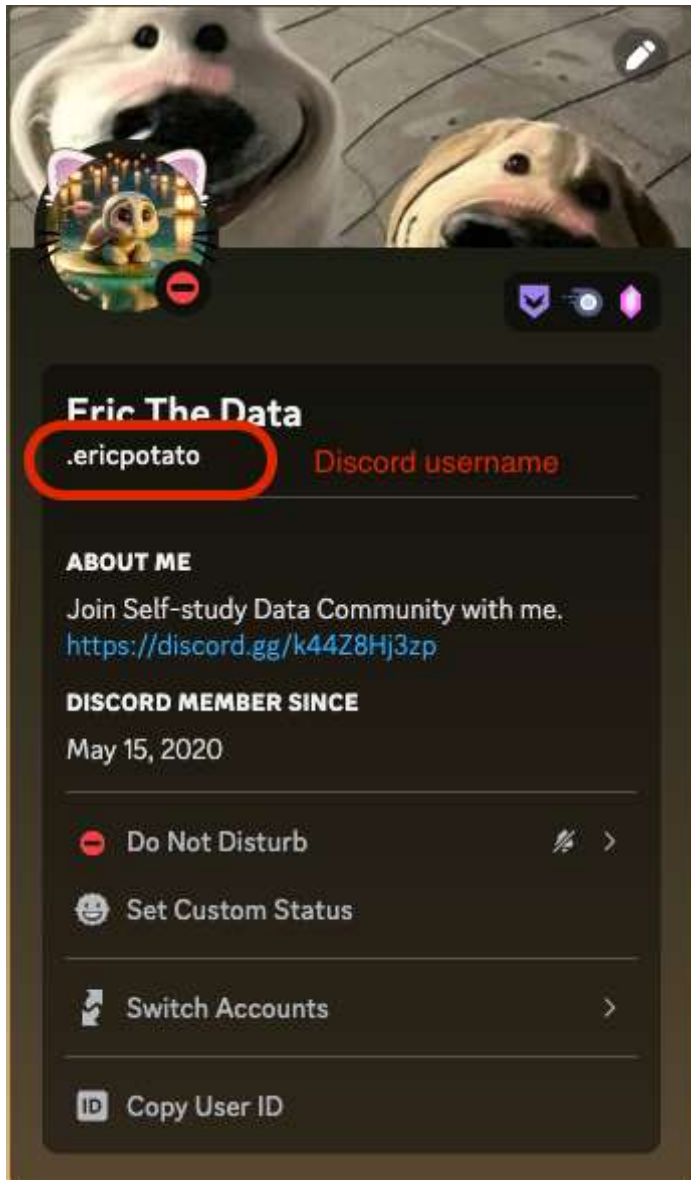
# MySQL Challenge Final Project

Tổng điểm 0/0 ?

- Please make sure you fill in your Discord username correctly, or your results won't be recorded.
- No uppercase characters or extra spaces.
- The deadline is 72 hours from the time I posted.
- For any questions, please post them in the channel #mysql-discussion.
- Please make sure to provide context or explanations to every queries as I cannot read your mind. (Note the result, for example how many rows returned)
- Save file as \*your username discord\*, only 1 SQL File (No zip, no docs...files)



## Discord Username



Discord username \*

Exercise Day \*

☒ Final Project

## Data Cleaning

### 1. **\*\*Data Consistency\*\***:

- Check for and correct inconsistencies.

### 2. **\*\*Removing Duplicates\*\***:

- Identify and remove duplicate rows if any.

**Hint:** UPDATE us\_household\_income\_clean  
SET `Type` = 'CDP'  
WHERE `Type` = 'CPD';

UPDATE us\_household\_income\_clean  
SET `Type` = 'Borough'  
WHERE `Type` = 'Boroughs';



## EDA

Here are 23 tasks to practice your MySQL skills and perform exploratory data analysis (EDA) to extract meaningful insights:

- Tasks 1 - 9: 0.5 points each.
- Tasks 10 - 22: 1 point each.
- Task 0: 2 point.

### Task 0: Summarizing Data by State (Special)

#### Question:

Assume that you will have new data each week. Can you please create store procedure and create event to active procedure every weeks to update and clean new data (From Cleaning Tasks)?

**Hint:** TimeStamp

### Task 1: Summarizing Data by State

#### Question:

Can you provide a summary report that shows the average land area and average water area for each state? Please include the state name and abbreviation, and order the results by the average land area in descending order.

### Task 2: Filtering Cities by Population Range

#### Question:

We need a list of all cities where the land area is between 50,000,000 and 100,000,000 square meters. Include the city name, state name, and county in the results, and order the list alphabetically by city name.



### Task 3: Counting Cities per State

#### Question:

Can you generate a report that counts the number of cities in each state? The report should include the state name and abbreviation and be ordered by the number of cities in descending order.

### Task 4: Identifying Counties with Significant Water Area

#### Question:

Please identify the top 10 counties with the highest total water area. The report should include the county name, state name, and total water area, ordered by total water area in descending order.

### Task 5: Finding Cities Near Specific Coordinates

#### Question:

We are looking for a list of all cities within a specific latitude and longitude range (latitude between 30 and 35, longitude between -90 and -85). Include the city name, state name, county, and coordinates, and order the results by latitude and then by longitude.

### Task 6: Using Window Functions for Ranking



**Question:**

We need to rank cities within each state based on their land area. Please use a window function to assign ranks and include the city name, state name, land area, and rank in your results. The report should be ordered by state name and rank.

**Task 7: Creating Aggregate Reports****Question:**

Can you generate a report showing the total land area and water area for each state, along with the number of cities in each state? Include the state name and abbreviation, and order the results by the total land area in descending order.

**Task 8: Subqueries for Detailed Analysis****Question:**

Can you provide a list of all cities where the land area is above the average land area of all cities? Use a subquery to calculate the average land area. The report should include the city name, state name, and land area, ordered by land area in descending order.

**Task 9: Identifying Cities with High Water to Land Ratios****Question:**

Can you identify cities where the water area is greater than 50% of the land area? Include the city name, state name, land area, water area, and the calculated water to land ratio, and order the results by the water to land ratio in descending order.



## Task 10: Dynamic SQL for Custom Reports

### Question:

Can you create a stored procedure that accepts a state abbreviation as input and returns a detailed report for that state? The report should include the total number of cities, average land area, average water area, and a list of all cities with their respective land and water areas.

## Task 11: Creating and Using Temporary Tables

### Question:

We need to create a temporary table that stores the top 20 cities by land area for further analysis. Use this temporary table to calculate the average water area of these top 20 cities, and include the city name, state name, land area, and water area in your final report.

## Task 12: Complex Multi-Level Subqueries

### Question:

Can you list all states where the average land area of cities is greater than the overall average land area across all cities in the dataset? Use multiple subqueries to calculate the overall average land area and the state-wise average land areas. Include the state name and average land area in your results.

## Task 13: Optimizing Indexes for Query Performance

### Question:

Can you analyze the impact of indexing on query performance? Create indexes on the columns State\_Name, City, and County, and compare the execution time of a complex query before and after indexing. Provide insights into how indexing improved (or did not improve) the query performance, including execution times and query plans.

## Task 14: Recursive Common Table Expressions (CTEs)

### Question:



Can you create a recursive CTE that calculates the cumulative land area for cities within each state, ordered by city name? Include the city name, state name, individual land area, and cumulative land area in your results.





## EDA

### Task 15: Data Anomalies Detection

#### Question:

Can you detect anomalies in the dataset, such as cities with unusually high or low land areas compared to the state average? Use statistical methods like Z-score or standard deviation to identify these anomalies. Include the city name, state name, land area, state average land area, and anomaly score in your results, ordered by the anomaly score in descending order.

### Task 16: Stored Procedures for Complex Calculations

#### Question:

Can you write a stored procedure that performs complex calculations, such as predicting future land and water area based on historical trends? The stored procedure should accept parameters for the city and state and return predicted values. Test the stored procedure with different inputs and document the results.

### Task 17: Implementing Triggers for Data Integrity

#### Question:

Can you write a trigger that automatically updates a summary table whenever new data is inserted, updated, or deleted in the main dataset? The summary table should include aggregated information such as total land area and water area by state. Test the trigger to ensure it correctly updates the summary table in response to changes in the main dataset.

### Task 18: Advanced Data Encryption and Security

#### Question:

We need to implement data encryption and ensure secure access to sensitive information in the dataset. Use MySQL encryption functions to encrypt columns such as Zip\_Code and Area\_Code. Demonstrate how to decrypt the data for authorized users and discuss the implications for data security.

### Task 19: Geospatial Analysis



**Question:**

We need to identify cities that fall within a specified radius from a given point (latitude and longitude). Include calculations to determine the distance between the given point and each city. Return the city name, state name, county, and calculated distance.

**Task 20: Analyzing Correlations****Question:**

Can you calculate the correlation between land area (ALand) and water area (AWater) for each state? Use statistical functions to determine the strength of the correlation. Include the state name and the correlation coefficient in your results.

**Task 21: Hotspot Detection****Question:**


Can you identify "hotspots" where the combination of land area and water area significantly deviates from the norm? Use statistical methods such as Z-scores or clustering to detect these hotspots. Include city name, state name, land area, water area, and the deviation score in your report.


**Task 22: Resource Allocation Optimization****Question:**

We need to optimize the allocation of resources (e.g., funding) based on the land and water area of each city. Create a model to distribute resources in a way that maximizes efficiency and equity. Include the city name, state name, land area, water area, and allocated resources in your results.



Final Project \*  
**Please send file that saved in ur Discord Username.**

 thocao811 - Mỹ ...

 Thêm tệp

Nội dung này không phải do Google tạo ra hay xác nhận. [Báo cáo](#) [Lạm dụng](#) - [Điều khoản Dịch vụ](#) - [Chính sách quyền riêng tư](#)

Google Biểu mẫu



